Hypothesis Testing with Interval Data: Case of Regulatory Constraints

Sa-aat Niwitpong
Department of Applied Statistics
King Mongkut's University of Technology North Bangkok
Bangkok 10800, Thailand
snw@kmitnb.ac.th

Hung T. Nguyen
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
hunguyen@nmsu.edu

Ingo Neumann
Geodetic Institute
Leibniz University of Hannover
D-30167 Hannover, Germany
neumann@gih.uni-hannover.de

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu

***Abstract*—In many practical situations, there exist regulatory thresholds: e.g., a concentration of certain chemicals in the car exhaust cannot exceed a certain level, etc. In this paper, we describe how to make accept/reject decisions under measurement or expert uncertainty in case of regulatory and expert-based thresholds – where the threshold does not come from a detailed statistical analysis.**

I. Hypothesis Testing: A General Problem

In many practical situations, it is desirable to check whether a given object (or situation) satisfies a given property. For example, we may want to check whether a patient has flu, whether a building or a bridge is structurally stable, etc.

In statistics, this problem is called *hypothesis testing:* we have a hypothesis – that a patient is healthy, that a building is structurally stable – and we want to test this hypothesis based on the available data. This hypothesis is usually called a *null hypothesis*, meaning that:

- if this hypothesis is satisfied then no ("null") action is required,
- while if this hypothesis is not satisfied, then we need to undertake some action: cure a patient, reinforce (or even evacuate) the structurally unstable building, etc.

II. Hypothesis Testing: Ideal Case of Complete Knowledge

Let us start with the idealized situation in which we have a complete knowledge about the object. In other words, we assume that we know the exact values of all the quantities $x_1, \ldots, x_n$ that characterize this object.

Since these quantities represent the complete knowledge about the object, this knowledge should be sufficient to determine whether the object satisfies the desired property. In other words,

- for some tuples $x = (x_1, \ldots, x_n)$ the corresponding objects satisfy the desired property, while
- for some other tuples x, the corresponding object does not satisfy the desired property.

Thus, the set X of all possible values of the tuple x can be divided into two subsets:

- the *acceptance region* consisting of all the tuples that satisfy the desired property; this region will be denoted by A; and
- the *rejection region* consisting of all the tuples that do not satisfy the desired property; this region will be denoted by R.

Thus, once we know the tuple x characterizing the given object, we:

- accept the hypothesis if $x \in A$, and
- reject the hypothesis if $x \in R$ (i.e., if $x \notin A$).

III. Hypothesis Testing: Realistic Case of Incomplete Knowledge

In practice, we usually only have an incomplete knowledge about an object: we only know the values of some of the parameters characterizing this objects, and even these values we only know approximately. Based on this partial information, we cannot always tell whether an object satisfies the given property, but there may be two objects which both lead to the same information such that one object satisfies the property and another one does not.

For example, if the property can be described as a restriction $x_1 + x_2 \le x_0$ on the sum of the two quantities x_1 and x_2, and we only know the value of the first quantity x_1, then:

- for some values x_1 (namely, for the values for which $x_2 \le x_0 - x_1$) we will have $x_1 + x_2 \le x_0$ and thus, the hypothesis is satisfied;
- on the other hand, for some other values x_1 (namely, for the values for which $x_2 > x_0 - x_1$) we will have $x_1 + x_2 > x_0$ and thus, the hypothesis is not satisfied.

In such situations, no matter what decision we make, this decision may turn to be erroneous. In general, there can be two types of errors:

- it is possible that the object actually satisfies the desired property, but we erroneously classify it as not satisfying the given null hypothesis; this error is called a *false positive*, or *Type I* error;
- it is also possible that the object actually does not satisfy the desired property, but we erroneously classify it as satisfying the given null hypothesis; this error is called a *false negative*, or *Type II* error.

IV. Traditional Statistical Approach to Hypothesis Testing

In the traditional statistical approach to hypothesis testing (see, e.g., [25]), we assume that we know the probability distribution of objects that satisfy the given hypothesis, and we are given the allowed probability p_0 of Type I error. In this situation, we select the accept and reject regions in such a way as to minimize the probability of Type II error.

For example, in a 1-D case, when we only know the value x_1 of a single quantity, the probability distribution is usually assumed to be Gaussian, with known mean a and know standard deviation σ. Usually, the anomalous situations correspond to too high (sometimes too low) values of this quantity: e.g., in medicine, the blood pressure is too high, or the cholesterol level, or the body temperature. In this case, we select the accept region as the set of all the values x_1 for which $x_1 \le x_0$ for some threshold x_0, and we select this threshold in such a way that the probability of exceeding this threshold is exactly p_0. For example:

- for $p_0 = 0.05 = 5\%$, we take $x = a + 2\sigma$;
- for $p_0 = 0.0005 = 0.05\%$, we take $x = a + 3\sigma$.

This traditional approach minimizes the probability of Type II errors. However, to find out what exactly is the probability of Type II error, we must also know the probability distribution of all the objects that do not satisfy the given hypothesis. This knowledge is also sometimes assumed in the traditional approach to hypothesis testing.

V. Limitations of the Traditional Statistical Approach to Decision Making and the Related Notion of Utility

Theoretically, the traditional approach is optimal. Theoretically, the traditional statistical approach to decision making is optimal – provided that we know the probability distribution of all the objects satisfying the given hypothesis and the threshold probability p_0.

Analysis of the assumptions behind the traditional statistical approach. In practice, we may not know the exact probability distribution, but the more "normal" (hypothesis-satisfying) objects we observe, the more accurately we can reconstruct this distribution.

The need to know the threshold probability p_0 is more serious, since there is no easy way to determine it empirically. From the common sense viewpoint, different values of p_0 correspond to different consequences of Type I errors.

Example. For example, in mass screening for breast cancer, hypothesis acceptance means that we consider the patient in good health and do not perform any further tests on this person. On the other hand, the rejection of null-hypothesis mean, in practice, that further, more expensive and more accurate tests need to be performed. In the ideal world, we should perform these additional tests on everyone, to minimize the probability of undetected breast cancer. In practice, our resources are limited, so we limit these additional tests to those patients for whom, based on the first screening, there is a reasonable possibility that these patients may have breast cancer.

In general, we must find the threshold probability p_0 from preferences. In general, the threshold probability is determined by the user possibilities and preferences. Thus, in practice, we must learn how to determine the threshold probability p_0 based on the user preferences. This determination is beyond the traditional statistical approach to hypothesis testing – an approach that assumes that this threshold probability p_0 is given. To determine p_0 from preferences, we must learn how to describe these preferences.

How to describe preferences: general idea. The possibility to describe preferences in precise terms comes from the fact that a decision maker can always decide which of the two alternatives is better (preferable). Thus, if we provide a continuous scale of alternatives, from a very bad to a very good one, then for each alternative in the middle, there should be an alternative on this scale which is, to this decision maker, equivalent to the given one.

How to describe preferences: specific ideas. Such a scale can be easy constructed as follows. We select two alternatives:

- a very negative alternative A_0; e.g., an alternative in which the decision maker loses all his money (and/or loses his health as well), and
- a very positive alternative A_1; e.g., an alternative in which the decision maker wins several million dollars.

Now, for every value $p \in [0, 1]$, we can consider a lottery in which we get A_1 with probability p and A_0 with the remaining probability $1 - p$. This probability will be denoted by $L(p)$.

For $p = 1$, the probability of the unfavorable outcome A_0 is 0, so the lottery $L(1)$ simply means the very positive alternative A_1. Similarly, for $p = 0$, the probability of the favorable outcome A_1 is 0, so the lottery $L(0)$ simply means the very negative alternative A_0. The larger the probability p, the more preferable the lottery $L(p)$. Thus, the corresponding lotteries $L(p)$ form a continuous 1-D scale ranging from the very negative alternative A_0 to the very positive alternative A_1.

The resulting notion of utility. Practical alternatives are usually better than $L(0) = A_0$ but worse than $L(1) = A_1$: $L(0) < A < L(1)$. Thus, for each practical alternative A, there exists a probability $p \in (0, 1)$ for which the lottery $L(p)$

is, to this decision maker, equivalent to A: $L(p) \sim A$. This "equivalent" probability p is called the *utility* of the alternative A and denoted by $u(A)$.

How can we actually find the value of this utility $u(A)$? We cannot just compare A with different lotteries $L(p)$ and wait until we get a lottery for which $L(p) \sim A$: there are many different probability values, so such a comparison would require an impractically long time. However, there is an alternative efficient way of determining $u(A)$ which is based on the following *bisection* procedure.

The main idea of this procedure is to produce narrower and narrower intervals containing the desired value $u(A)$. In the beginning, we only know that $u(A) \in [0, 1]$, i.e., we know that $u(A) \in [\underline{u}, \overline{u}]$ with $\underline{u} = 0$ and $\overline{u} = 1$. Let us assume that at some iteration of this procedure, we know that $u(A) \in [\underline{u}, \overline{u}]$, i.e., that $L(\underline{u}) \le A \le L(\overline{u})$. To get a narrower interval, let us take the midpoint $m \stackrel{\text{def}}{=} \dfrac{\underline{u} + \overline{u}}{2}$ of the existing interval and compare $L(m)$ with A.

- If A is better than $L(m)$ ($L(m) \le A$), this means that $m \le u(A)$ and thus, that the utility $u(A)$ belongs to the upper half-interval $[m, \overline{u}]$.
- If A is worse than $L(m)$ ($A \le L(m)$), this means that $u(A) \le m$ and thus, that the utility $u(A)$ belongs to the lower half-interval $[\underline{u}, m]$.

In both cases, we get a new interval containing $u(A)$ whose width is the half of the width of the interval $[\underline{u}, \overline{u}]$. We start with an interval of width 1. Thus, after k iterations, we get an interval $[\underline{u}, \overline{u}]$ of width 2^{-k} that contains $u(A)$. In this case, both endpoints $\underline{u}$ and $\overline{u}$ are 2^{-k}-approximations to $u(A)$. In particular:

- to obtain $u(A)$ with accuracy $1\% = 0.01$, it is sufficient to perform 7 iterations: since $2^{-7} = 1/128 < 0.01$;
- to obtain $u(A)$ with accuracy $0.1\% = 0.001$, it is sufficient to perform 10 iterations: since
$$2^{-10} = 1/1024 < 0.001;$$
- to obtain $u(A)$ with accuracy $10^{-4}\% = 10^{-6}$, it is sufficient to perform 20 iterations: since
$$2^{-20} = 1/(1024)^2 < 10^{-6}.$$

The numerical value of the utility depends on the choice of extreme alternatives A_0 and A_1. In our definition, the numerical value of the utility depends on the selection of the alternatives A_0 and A_1: e.g., A_0 is the alternative whose utility is 0 and A_1 is the alternative whose utility is 1. What if we use a different set of alternatives, e.g., $A_0' < A_0$ and $A_1' > A_1$?

Let A be an arbitrary alternative between A_0 and A_1, and let $u(A)$ be its utility with respect to A_0 and A_1. In other words, we assume that A is equivalent to the lottery in which we have

- A_1 with probability $u(A)$ and
- A_0 with probability $1 - p$.

In the scale defined by the new alternatives A_0' and A_1', let $u'(A_0)$, $u'(A_1)$, and $u'(A)$ denote the utilities of A_0, A_1, and A. This means, in particular, that

- A_0 is equivalent to the lottery in which we get A_1' with probability $u'(A_0)$ and A_0' with probability $1 - u'(A_0)$; and
- A_1 is equivalent to the lottery in which we get A_1' with probability $u'(A_1)$ and A_0' with probability $1 - u'(A_1)$.

Thus, the alternative A is equivalent to the compound lottery, in which

- first, we select A_1 or A_0 with probabilities $u(A)$ and $1 - u(A)$, and then
- depending on the first selection, we select A_1' with probability $u'(A_1)$ or $u'(A_0)$ – and A_0' with the remaining probability.

As the result of this compound lottery, we get either A_0' or A_1'. The probability p of getting A_1' in this compound lottery can be computed by using the formula of full probability

$$p = u(A) \cdot u'(A_1) + (1 - u(A)) \cdot u'(A_0) =$$
$$u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0).$$

So, the alternative A is equivalent to a lottery in which we get A_1' with probability p and A_0' with the remaining probability $1 - p$. By definition of utility, this means that the utility $u'(A)$ of the alternative A in the scale defined by A_0' and A_1' is equal to this value p:

$$u'(A) = u(A) \cdot (u'(A_1) - u'(A_0)) + u'(A_0).$$

So, changing the scale means a linear re-scaling of the utility values:

$$u(A) \to u'(A) = a \cdot u(A) + b$$

for some $a = u'(A_1) - u'(A_0) > 0$ and $b = u'(A_0)$.

Vice versa, for every $a > 0$ and b, one can find appropriate events A_0' and A_1' for which the re-scaling has exactly these values a and b. In other words, utility is defined modulo an arbitrary (increasing) linear transformation.

Utility of an action: a derivation of the expected utility formula. What if an action leads to alternatives $a_1, \ldots, a_m$ with probabilities $p_1, \ldots, p_m$? Suppose that we know the utility $u_i = u(a_i)$ of each of the alternatives $a_1, \ldots, a_m$. By definition of the utility, this means that for each i, the alternative a_i is equivalent to the lottery $L(u_i)$ in which we get A_1 with probability u_i and a_i with probability $1 - u_i$. Thus, the results of the action are equivalent to the "compound lottery" in which, with the probability p_i, we select a lottery $L(u_i)$. In this compound lottery, the results are either A_1 or A_0. The probability p of getting A_1 in this compound lottery can be computed by using the formula for full probability:

$$p = p_1 \cdot u_1 + \ldots + p_m \cdot u_m.$$

Thus, the action is equivalent to a lottery in which we get A_1 with probability p and A_0 with the remaining probability $1 - p$. By definition of utility, this means that the utility u of the action in question is equal to

$$u = p_1 \cdot u_1 + \ldots + p_m \cdot u_m.$$

In statistics, the right-hand of this formula is known as the *expected value*. Thus, we can conclude that the utility of each action with different possible alternatives is equal to the expected value of the utility; see, e.g., [11], [16], [24].

VI. Previous Work: What If the Object's Characteristics Are Only Known with Uncertainty

In the traditional statistical approach – and in its utility-based extension – it is usually assumed that

- we know the probability distribution for all the objects that satisfy the given property exact values of the parameters, and
- we know the exact value of the characteristic(s) that describe a given object.

As we have mentioned, in practice, we only have a partial knowledge about the corresponding probability distribution, but the more objects we observe, the more accurately we can determine this distribution. Thus, in many practical situations, it is reasonable to ignore the corresponding approximation error and assume that we know the exact probability distribution.

The value characterizing an object usually comes from measurement and/or from expert estimates and is therefore also only known with uncertainty. However, no matter how many objects we have observed, this uncertainty does not decrease. Thus, it is less reasonable to ignore this uncertainty.

In our survey paper [13], we described how the traditional statistical approach (and its utility-based extension) can be modified if we take this uncertainty into account; see also [1], [3], [4], [6], [14], [15], [17], [18], [19]. The results depend on what we know about the corresponding uncertainty.

VII. Types of Uncertainty: Probabilistic, Interval, Fuzzy

Uncertainty means that the estimate $\widetilde{x}$ that we obtained from measurements or from experts is, in general, different from the actual (unknown) value x of the estimated quantity. In other words, uncertainty means that, in general, we have a non-zero approximation error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$.

In the ideal case, we know the probabilities of different possible values of approximation error Δx. This is an assumption behind the traditional textbook approach to handling measurement errors. The corresponding situation is called *probabilistic uncertainty*; see, e.g., [23].

In many practical cases, however, we do not know the probabilities, we only know the upper bound Δ on the (absolute value of) the approximation error: $|\Delta x| \le \Delta$. In this case, based on the approximate value $\widetilde{x}$, we can conclude that the actual (unknown) value of the quantity x is somewhere in the interval $\mathbf{x} \stackrel{\text{def}}{=} [\widetilde{x} - \Delta, \widetilde{x} + \Delta]$ [23]. This situation is called *interval uncertainty* [8], [9].

In addition to the guaranteed (no uncertainty) bound Δ on the approximation error Δx, experts can usually provide us with smaller bounds corresponding to different degrees of uncertainty. In this case, instead of a single interval $\mathbf{x}$, we have different intervals corresponding to different levels of uncertainty.

We can gauge the expert's uncertainty as a number α between 0 and 1 (0 means no uncertainty at all). So, we have intervals $\mathbf{x}(\alpha)$ corresponding to different levels of uncertainty.

The narrower the interval, the less we are certain about this interval. So, for each $\alpha < \alpha'$, the interval $\mathbf{x}(\alpha')$ is a subinterval of the interval $\mathbf{x}(\alpha)$. In this sense, we have a *nested* family of intervals. This family of intervals can be alternatively described if for each value x, we describe the largest value of uncertainty α for which $x \in \mathbf{x}(\alpha)$; the function μ that maps a value x into this largest value is called a *membership function*, or a *fuzzy set*; see, e.g., [12], [21], [22]. Vice versa, once we have a fuzzy set $\mu(x)$, we can determine the intervals $\mathbf{x}(\alpha)$ as *α-cuts* $\{x : \mu(x) \ge \alpha\}$. This situation is called *fuzzy uncertainty*.

VIII. A New Important Case of Hypothesis Testing: Testing Whether an Object Satisfies Given Regulations

In the previous text, we considered situations when we know the probability distribution of objects that satisfy the null hypothesis. There is a practically important case of hypothesis testing when we do not know these probabilities: the cases of regulatory thresholds.

In many practical situations, we are given regulatory thresholds such as "the speed limit is 75 miles", "the length of the machine axis has to be between 13.2 m and 13.21 m", or "a concentration of certain chemicals in the car exhaust cannot exceed a certain level", etc.

In general, we have:

- the acceptance region A consisting of all the values that satisfy given regulations, and
- the rejection region R consisting of all the values that do not satisfy the regulations.

Our objective is to check whether the given object satisfies the corresponding regulations, i.e., whether it belongs to the acceptance region A or to rejection region R.

IX. Ideal Case: We Know the Exact Value of the Tested Quantity

When we know the exact value x of the regulated quantity, then testing the corresponding regulation is easy: we just check whether this value x belongs to the acceptance region A.

For example, in case of a single regulatory inequality, we simply compare the value x with a threshold x_0:

- if the actual value x is within the required bounds, i.e., if $x \le x_0$, then we conclude that the regulations are satisfied;
- on the other hand, if the actual value exceed the desired threshold, i.e., if $x > x_0$, then the corresponding regulation is violated.

X. Practical Situation: We Only Know the Approximate Value of the Tested Quantity

In practice, the situation is more complicated. The values of the desired quantities come from measurements or from expert estimates. The estimate $\widetilde{x}$ obtained from a measurement or from an expert estimate is never absolutely accurate. In other words, the estimate $\widetilde{x}$ is, in general, somewhat different from the actual (unknown) value x of the desired quantity. It is therefore necessary to be able, given this estimate $\widetilde{x}$, to tell

- whether the actual (unknown) value x is acceptable (below the threshold), or
- whether the actual (unknown) value x is not acceptable (above the threshold).

This is the problem that we will be handling in the present paper.

The main difficulty in solving the above problem is that our estimate $\widetilde{x}$ about the desired quantity x is not absolutely accurate. So, to formulate the regulatory case of the hypothesis testing problem in exact terms, we must describe what information we have about the accuracy of our estimate. Thus, we need to consider cases of probabilistic, interval, or fuzzy uncertainty.

XI. Case of Probabilistic Uncertainty

What is probabilistic uncertainty: reminder. In the case of probabilistic uncertainty, we assume that we know the probability distribution of the approximation error Δx.

How to describe probabilistic uncertainty. A probability distribution is usually characterized

- either by a probability density function $\rho(t)$,
- or by a cumulative distribution function (cdf) $F(t) \stackrel{\text{def}}{=} \text{Prob}(\Delta x \le t)$.

Under probabilistic uncertainty, we are not 100% certain about the validity of the hypothesis. In general, for each estimate $\widetilde{x}$ of the quantity x, it is possible that the actual (unknown) value $x = \widetilde{x} - \Delta x$ belongs to the acceptance region A, and it is also possible that the value x belongs to the rejection region.

Estimating probability of satisfying the hypothesis: general case. Since we know the probabilities of different possible values of the approximation error $\Delta x = \widetilde{x} - x$, we can compute the probability p_A that the actual value x is acceptable as the overall probability that $x \in A$, i.e., as

$$p_A = \int_A \rho(\widetilde{x} - x)\, dx.$$

The probability p_R that the object does not satisfy the hypothesis is thus equal to

$$p_R = 1 - p_A.$$

Estimating probability of satisfying the hypothesis: important practical cases. In particular, if the acceptance region is determined by a threshold, i.e., if $A = \{x : x \le x_0\}$, then p_A is the probability that $x \le x_0$, i.e., in terms of the approximation error $\Delta x = \widetilde{x} - x$, the probability that $\Delta x \ge \widetilde{x} - x_0$. The probability of Δx being $\le \widetilde{x} - x_0$ is, by definition, equal to $F(\widetilde{x} - x_0)$. Thus, the probability of the opposite event is equal to

$$p_A = 1 - F(\widetilde{x} - x_0).$$

(We consider the case when the probability distribution has a density; in this case, the probability that Δx is exactly equal to $\widetilde{x} - x_0$ is 0.)

If the acceptance region is determined by a lower threshold l_0, i.e., if $A = \{x : x \ge l_0\}$, then, similarly, p_A is the probability that $l_0 \le x$, i.e., in terms of $\Delta x = \widetilde{x} - x$, the probability that $\Delta x \le \widetilde{x} - l_0$. The probability of Δx being $\le \widetilde{x} - l_0$ is, by definition, equal to $F(\widetilde{x} - l_0)$. Thus, we have

$$p_A = F(\widetilde{x} - l_0).$$

If the acceptance region is an interval

$$A = \{x : l_0 \le x \le x_0\},$$

then p_A is the probability that $l_0 \le x \le x_0$, i.e., in terms of $\Delta x = \widetilde{x} - x$, the probability that $\widetilde{x} - x_0 \le \Delta x \le \widetilde{x} - l_0$. This probability can be computed as the difference between the probability that $\Delta x \le \widetilde{x} - l_0$ and the probability that $\Delta x \le \widetilde{x} - x_0$, i.e., as

$$p_A = F(\widetilde{x} - l_0) - F(\widetilde{x} - x_0).$$

Relation to Type I and Type II errors. If we classify the object with estimated value $\widetilde{x}$ as satisfying the hypothesis, then the probability p_R that it actually does not satisfy the regulations is the probability of the Type II error.

Similarly, if we classify the object with estimated value $\widetilde{x}$ as not satisfying the hypothesis, then the probability p_A that it actually does satisfy the regulations is the probability of the Type I error.

How to decide whether the hypothesis is satisfied: utility approach. In order to decide whether to classify the object as satisfying or not satisfying, we must know the consequences of each type of error. As we have already mentioned, these consequences can be described in terms of utilities; then we should make a decision that leads to the largest value of expected utility.

So, to make a decision, we must know the utilities corresponding to all possible situations:

- the utility u_{++} of the situation in which the actual value is acceptable, and we (correctly) classify it as acceptable;
- the utility u_{+-} of the situation in which the actual value is acceptable, but we (incorrectly) classify it as not acceptable;
- the utility u_{-+} of the situation in which the actual value is not acceptable, but we (incorrectly) classify it as acceptable;

- the utility u_{--} of the situation in which the actual value is not acceptable, and we (correctly) classify it as not acceptable.

Correct decisions are preferable to incorrect ones, so we should have $u_{++} > u_{+-}$ and $u_{--} > u_{-+}$.

The actual object with the estimated value $\widetilde{x}$ is acceptable with probability p_A and not acceptable with the probability $p_R = 1 - p_A$. Hence, the expected utility u_A of the acceptance decision is

$$u_A = p_A \cdot u_{++} + (1 - p_A) \cdot u_{-+},$$

and the expected utility u_R of the rejection decision is

$$u_R = p_A \cdot u_{+-} + (1 - p_A) \cdot u_{--}.$$

We thus select the acceptance decision when $u_A \geq u_R$, i.e., when

$$p_A \cdot u_{++} + (1 - p_A) \cdot u_{-+} \geq p_A \cdot u_{+-} + (1 - p_A) \cdot u_{--}.$$

Moving terms proportional to p_A to the left and all other terms to the right, we conclude that

$$p_A \cdot (u_{++} - u_{-+} - u_{+-} + u_{--}) \geq u_{--} - u_{-+}.$$

Since $u_{++} > u_{+-}$ and $u_{--} > u_{-+}$, we have

$$u_{++} - u_{-+} - u_{+-} + u_{--} =$$
$$(u_{++} - u_{+-}) + (u_{--} - u_{-+}) > 0.$$

Dividing both sides of the above inequality by this positive number, we get the following final criterion.

Resulting criterion: general case. We accept the hypothesis if

$$p_A \geq p^{(0)},$$

where

$$p^{(0)} \stackrel{\text{def}}{=} \frac{u_{--} - u_{-+}}{u_{++} - u_{-+} - u_{+-} + u_{--}}.$$

Resulting criterion: important practical cases. If the acceptance region is determined by a threshold $A = \{x : x \leq x_0\}$, then $p_A = 1 - F(\widetilde{x} - x_0)$ and thus, the above inequality is equivalent to $F(\widetilde{x} - x_0) \leq 1 - p^{(0)}$. Since the cdf $F(t)$ is an increasing function, this is equivalent to $\widetilde{x} - x_0 \leq F^{-1}(1 - p^{(0)})$, where F^{-1} denotes the inverse function to cdf. Thus, in this case, we accept the hypothesis if

$$\widetilde{x} \leq x_0 + F^{-1}(1 - p^{(0)}).$$

It is worth mentioning that if $F^{-1}(1 - p^{(0)}) > 0$ and $x_0 < \widetilde{x} < x_0 + F^{-1}(1 - p^{(0)})$, then, based on the fact that the estimate $\widetilde{x}$ exceeds x_0, one may be tempted to classify this object as unacceptable; however, due to the uncertainty, the actual value x may be different, in particular, it may be smaller than x_0. To avoid the corresponding Type I errors, we classify this object as acceptable.

Similarly, if the acceptance region is determined by a lower threshold $A = \{x : x \geq l_0\}$, then $p_A = F(\widetilde{x} - l_0)$ and thus, the above inequality is equivalent to $F(\widetilde{x} - l_0) \geq p^{(0)}$. Since the cdf $F(t)$ is an increasing function, this is equivalent to $\widetilde{x} - l_0 \geq F^{-1}(p^{(0)})$. Thus, in this case, we accept the hypothesis if

$$\widetilde{x} \geq l_0 + F^{-1}(p^{(0)}).$$

Here, similarly, if $F^{-1}(p^{(0)}) > 0$ and $l_0 < \widetilde{x} < l_0 + F^{-1}(p^{(0)})$, then, based on the fact that the estimate $\widetilde{x}$ exceeds l_0, one may be tempted to classify this object as acceptable; however, due to the uncertainty, the actual value x may be different, in particular, it may be smaller than l_0. To avoid the corresponding Type II errors, we classify this object as unacceptable.

If the acceptance region is an interval

$$A = \{x : l_0 \leq x \leq x_0\},$$

then $p_A = F(\widetilde{x} - l_0) - F(\widetilde{x} - x_0)$ and thus, the criterion for concluding that the object with estimate $\widetilde{x}$ satisfies the given hypothesis is

$$F(\widetilde{x} - l_0) - F(\widetilde{x} - x_0) \geq p^{(0)}.$$

Illustrative example. As an illustrative example, let us consider car testing for exhaust pollution. We need to check that the level of each potential pollutant (hydrocarbons, carbon monoxide, nitrogen oxide) does not exceed the maximum permissible level x_0. The accuracy of this testing is about 15-20%. In the probabilistic approach, it is therefore reasonable to assume that the measurement error $\widetilde{x} - x$ is normally distributed with 0 mean and standard deviation $\sigma = 0.175x_0$. We need to decide whether to accept the hypothesis that the car is not polluting (then we will issue, to this car, an annual state certification allowing it to drive), or to conclude that the exhaust pollution is (probably) excessive and therefore, the car exhaust system must be re-tuned before the car is allowed on the road.

In the El Paso region of the State of Texas, the average cost of a car exhaust tuning is about \$60. The cost of a polluting car to to the environment can be estimated based on the fact that the State of Texas offers a \$3000 voucher to every driver who wants to trade in an older more polluting car for a newer better model (and whose income is below a certain threshold). So, in this case, the cost of a "reject" decision is simply the cost of tuning $u_{--} = u_{+-} = -60$, the cost of an "accept" decision for a polluting car is $u_{-+} = -3000$, and the cost of an "accept" decision for a non-polluting car is $u_{++} = 0$. In this case, $p^{(0)} = 2940/3000 \approx 0.98$, so $F^{-1}(1 - p^{(0)}) = F^{-1}(0.02) \approx -2.3\sigma = -2.3 \cdot 0.175x_0 \approx -0.4x_0$. Thus, we should decide that the car passed the inspection if $\widetilde{x} \leq x_0 + (-0.4x_0) = 0.6x_0$.

Please note that here, the acceptance threshold is very low, 0.6 of the nominal value. The reason for this lowness is that the cost of a Type I error is very small in comparison with the cost of a Type II error; as a result, we tend to err on the side of requiring good cars to be re-tuned.

XII. Case of Interval Uncertainty

Description of the situation. As we have mentioned earlier, in many practical situations, we do not know the probabilities of different values of the approximation error Δx, we only know the upper bound on this error. In this situation, the only information that we have about the (unknown) actual value x of the desired quantity is that this value x belongs to the interval $\mathbf{x} = [\underline{x}, \overline{x}]$.

Under such interval uncertainty, we must decide whether to accept or to reject the null hypothesis.

Simple cases. There are two cases, when the classification under interval uncertainty is easy:

- If $\mathbf{x} \cap A = \emptyset$, this means that all possible values of x belong to the rejection region R. In this case, we know that the corresponding object belongs to the rejection region (i.e., does not satisfy the null hypothesis).
- If $\mathbf{x} \cap R = \emptyset$, this means that all possible values of x belong to the acceptance region A. In this case, we know that the corresponding object belongs to the acceptance region (i.e., satisfies the null hypothesis).

Remaining case: description. The remaining case is when the interval $\mathbf{x}$ of possible values of x contains both values which are acceptable and values which are not acceptable.

Remaining case: utility. To make a decision, we must consider utility associated with both acceptance and rejection decisions.

In the remaining case, the utility depends on whether the actual value is acceptable or not.

- If we make an acceptance decision, then the utility is u_{++} if the object is actually acceptable and u_{-+} ($< u_{++}$) if the object is not acceptable.
- If we make a rejection decision, then the utility is u_{--} if the object is actually not acceptable and u_{+-} ($< u_{--}$) if the object is acceptable.

In both cases, instead of a single value u of utility, we have a *pair* $\{\underline{u}, \overline{u}\}$ of possible values. How can we then make decisions based on this pair?

Main idea underlying utility theory: reminder. The main idea behind utility theory is that to gauge the quality of each situation, we describe it by a single utility value. In line with this general idea, to gauge the quality of a situation described by a pair $\{\underline{u}, \overline{u}\}$, we should find a utility value u which is (in some reasonable sense) equivalent to this pair.

Additional idea: invariance. Our objective is to develop a mapping $e(\underline{u}, \overline{u})$ that maps every pair $\{\underline{u}, \overline{u}\}$ into a single equivalent value $u = e(\underline{u}, \overline{u})$. What properties should this mapping have?

As we have mentioned, the numerical values of the utility depend on the choice of the two extreme alternatives A_0 and A_1. Different choices of these two extreme alternatives lead to different scales for representing utility. Different scales $u(A)$ and $u'(A)$ are related to each other by a linear transformations $u'(A) = a \cdot u(A) + b$ for some $a > 0$ and b.

It is therefore reasonable to require that the desired mapping does not change under such re-scalings. Let us formulate this property in precise terms. Suppose that we start in the original scale. In this case, we have a pair $\{\underline{u}, \overline{u}\}$. Based on this pair, we find the equivalent value $u = e(\underline{u}, \overline{u})$.

Suppose now that we use a different scale to represent the same situation, a scale which is related to the original one by a linear transformation $u'(A) = a \cdot u(A) + b$. In this new scale, the elements $\underline{u}$ and $\overline{u}$ of the pair take new numerical values $\underline{u}' = a \cdot \underline{u} + b$ and $\overline{u}' = a \cdot \overline{u} + b$. When we apply the combination function e to these new values $\underline{u}'$ and $\overline{u}'$, we get an equivalent value $u' = e(\underline{u}', \overline{u}')$, i.e.,

$$u' = e(a \cdot \underline{u} + b, a \cdot \overline{u} + b).$$

It is reasonable to require that this new value represent the exact same equivalent utility u as before, but expressed in the new scale, i.e., that $u' = a \cdot u + b$ for $u = e(\underline{u}, \overline{u})$.

Substituting the expressions $u' = e(a \cdot \underline{u} + b, a \cdot \overline{u} + b)$ and $u = e(\underline{u}, \overline{u})$ into the formula $u' = a \cdot u + b$, we conclude that for every $\underline{u} < \overline{u}$, $a > 0$, and b, we have

$$e(a \cdot \underline{u} + b, a \cdot \overline{u} + b) = a \cdot e(\underline{u}, \overline{u}) + b.$$

Let us show that this natural invariance condition leads to a very specific expression for the combination function u.

Consequences of invariance. Let us pick one possible pair, e.g., a pair $\{0, 1\}$. This means that the actual utility of a situation can be either 0 and 1, depending on the circumstances.

Let us denote the utility value $e(0, 1)$ equivalent to this pair by α. From the common sense viewpoint, this value should be between 0 and 1: $\alpha \in [0, 1]$.

Let $\{u^-, u^+\}$ be an arbitrary non-degenerate pair. One can easily check that this pair can be obtained from the pair $[0, 1]$ by an appropriate linear re-scaling: namely, from the conditions that $\{a \cdot 0 + b, a \cdot 1 + b\} = \{u^-, u^+\}$ we conclude that $a \cdot 0 + b = b = u^-$. Then, from $a \cdot 1 + b = a + b = u^+$, we conclude that $a = u^+ - b = u^+ - u^-$. For the resulting values $\underline{u} = 0$, $\overline{u} = 1$, $a = u^+ - u^-$, and $b = u^-$, the above invariance implies that $e(u^-, u^+) = (u^+ - u^-) \cdot \alpha + u^-$. By combining terms proportional to u^- and to u^+, we conclude that

$$u = \alpha \cdot u^+ + (1 - \alpha) \cdot u^-.$$

This is exactly the formula originally proposed by the Nobelist by L. Hurwicz [7]. So, we arrive at the following solution to the problem of hypothesis testing under interval uncertainty:

Hurwicz-type solution to hypothesis testing under interval uncertainty. We characterize this situation (decision) by a single equivalent utility value

$$u = \alpha \cdot \overline{u} + (1 - \alpha) \cdot \underline{u},$$

and we select a decision for which the equivalent value u is the largest [7].

Specifically, we classify the object as satisfying the hypothesis if

$$\alpha \cdot u_{++} + (1 - \alpha) \cdot u_{-+} \geq \alpha \cdot u_{--} + (1 - \alpha) \cdot u_{+-},$$

and classify the object as not satisfying the hypothesis if

$$\alpha \cdot u_{++} + (1-\alpha) \cdot u_{-+} < \alpha \cdot u_{--} + (1-\alpha) \cdot u_{+-}.$$

In other words, we classify the object as satisfying the hypothesis if

$$\alpha \cdot (u_{++} - u_{-+} - u_{+-} + u_{--}) \geq u_{--} - u_{-+},$$

i.e.,

$$\alpha \geq p^{(0)},$$

where

$$p^{(0)} = \frac{u_{--} - u_{-+}}{u_{++} - u_{-+} - u_{+-} + u_{--}}.$$

How do we select α: Hurwicz's interpretation. The above approach requires that we fix the value of the parameter α. This parameter must be selected in such a way as to best represent the user's preferences. To help with this selection, L. Hurwicz provided the following reasonable interpretation of this parameter.

Let us recall that in case of the interval uncertainty, we do not know the exact value of the utility characterizing each decision, we only know the pair $\{\underline{u}, \overline{u}\}$ of possible values of this utility.

- In the most optimistic case, we get the largest possible value $\overline{u}$ of this utility.
- In the most pessimistic case, we get the smallest possible value $\underline{u}$ of this utility.

It turns out that these cases are directly related to the choice of the parameter α:

- When $\alpha = 1$, this means the equivalent utility value is equal to $u = \overline{u}$. In other words, we judge each decision by it most optimistic outcome.
- When $\alpha = 1$, this means the equivalent utility value is equal to $u = \underline{u}$. In other words, we judge each decision by it most pessimistic outcome.
- When $0 < \alpha < 1$, this means the equivalent utility value u is strictly in between the pessimistic value $\underline{u}$ and the optimistic value $\overline{u}$.

In view of this relation, the general Hurwicz criterion for decision making under interval uncertainty is also called *optimism-pessimism criterion* – because to make a decision, it uses a linear combination of the optimistic and pessimistic estimates.

In case of regulatory and expert-based thresholds, after fixing the parameter α, we have a clear algorithm for hypothesis testing.

Illustrative example. Let us illustrate the above approach on the above example of pollution testing for a car exhaust. In the interval approach, it is reasonable to interpret the reported 15-20% measurement accuracy as the $0.2x_0$ upper bound the (absolute value of) the measurement error. In other words, when the measurement result is $\widetilde{x}$, we assume that the actual pollution level can take any value from the interval $[\widetilde{x} - 0.2x_0, \widetilde{x} + 0.2x_0]$.

In this case, if $\widetilde{x} + 0.2x_0 \leq x_0$, i.e., if $\widetilde{x} \leq 0.8x_0$, then we are absolutely sure that the actual pollution value is below the maximum allowed level x_0 and thus, the car should be certified as driveable.

Similarly, if $\widetilde{x}_0 - 0.2x_0 > x_0$, i.e., if $\widetilde{x} > 1.2x_0$, then we are absolutely sure that the actual pollution value is above the maximum allowed level x_0 and thus, the car should be re-tuned.

In the remaining cases, when $0.8x_0 < \widetilde{x} \leq 1.2x_0$, we must use the Hurwicz-type solution. In this example, the utility values related to different solutions lead to $p^{(0)} \approx 0.98$. Thus, unless we are extremely optimistic ($\alpha > 0.98$), we should reject the hypothesis and request the car to be re-tuned.

Limitations of Hurwicz approach. One disadvantage of this approach is the fact that (after fixing the parameter α) wherever the interval $\mathbf{x}$ is located, as long as even a minimal part of the interval is inside the acceptance region and even a minimal part of the interval is inside the rejection region, we have the exact same decision. In other words, we have the same decision when most of the interval is in the acceptance region and when most of the interval $\mathbf{x}$ is in the rejection region.

An alternative approach. An alternative approach is to assume that there exists a probability distribution inside the interval. A reasonable assumption is that the distribution inside the interval $\mathbf{x}$ is uniform [14], [18], because it is the maximum entropy distribution among all continuous distributions which are supported in the interval, see, e.g., [10]. In this case, the probability to be within the acceptance region is equal to the ratio

$$p = \frac{|\mathbf{x} \cap A|}{|\mathbf{x}|},$$

where $|\mathbf{x}|$ denotes the width of the interval $\mathbf{x}$.

Then, we can use the criterion derived for the probabilistic case, and accept the null hypothesis if this probability exceeds the above-derived threshold $p^{(0)}$: $p \geq p^{(0)}$.

Illustrative example. In the car exhaust example, if we have $\widetilde{x} = 0.801x_0$, this meas that the actual value x of the pollution can be anywhere between $0.601x_0$ and $1.001x_0$. Our objective is to make sure that $x \leq x_0$. In the interval $[0.601x_0, 1.001x_0]$, only a tiny portion is above x_0, so we should expect this car to pass the test. However, according to the above Hurwicz-type solution, this car should be re-tuned.

Let us check that the alternative idea indeed leads to more intuitively plausible decision. Indeed, in this approach, the width $|\mathbf{x}|$ of the interval $\mathbf{x} = [0.601x_0, 1.001x_0]$ is equal to $0.4x_0$, while the width $|\mathbf{x} \cap A|$ of the intersection $\mathbf{x} \cap A = [0.601x_0, 1.001x_0] \cap [0, x_0] = [0.601x_0, x_0]$ is equal to $0.399x_0$. Thus, the ratio $p = \dfrac{|\mathbf{x} \cap A|}{|\mathbf{x}|}$ is equal to $p = 0.399/0.4 = 0.9975$. Since $p_0 = 0.98$, we have $p \geq 0.98$, thus, we do consider this car to be acceptable.

XIII. Case of Fuzzy Uncertainty

A fuzzy set is a natural next step after an interval: reminder. As we have mentioned earlier, a natural way to view a fuzzy set is to view it as a natural generalization of the notion of an interval – namely, as a nested family of intervals. Because of this view, in order to solve a hypothesis testing problem under fuzzy uncertainty, it is reasonable

- to recall how this problem is solved under interval uncertainty, and
- to generalize to the case of fuzzy uncertainty.

Case of interval uncertainty: reminder. For interval uncertainty, we described the probability p_A that a quantity described by the interval $X = \mathbf{x}$ is acceptable as the conditional probability that, given a random value from this interval X, we get an element from the accept region A. Then, we concluded that a value described by an interval X is acceptable if p_A exceeds the threshold $p^{(0)}$.

Formally, the corresponding conditional probability can be described as $p_A = P(A \cap X \mid X)$. By definition of conditional probability, we can describe this probability as

$$p_A = \frac{P(A \cap X)}{P(X)}.$$

In the interval case, to get the probabilities, we assumed that the values x are uniformly distributed on the interval X. In order to extend this idea to the fuzzy case, we must extend this assumption to the fuzzy case. For that extension, it is useful to recall the known relationship between fuzzy and probabilities.

Known relationship between fuzzy and probabilities: a reasonable way to gauge the degree of certainty. In our description of fuzzy uncertainty, we did not specify how to gauge the degree of uncertainty – and, correspondingly, how to gauge the values $\mu(x)$ of the membership function. There exist many such gauging schemes; see, e.g., [12], [22].

One reasonable way to do it is to poll experts and to select, as $\mu(x)$, the proportion of experts who believe that the value x is indeed possible. In this case, $\mu(x)$ is a probability that a randomly selected expert believes that the value x is possible.

Relation with random sets. The above description can be reformulated in more mathematical term. Every expert has a set of values that this experts considers possible. We consider the experts to be equally probable, so these sets are equally probable.

Thus, we have, in effect, a probability distribution on the class of all possible sets. Similarly to the fact that the probability distribution on the set of all possible numbers is called a random number, the probability distribution on the class of all possible sets is called a *random set.* Thus, a membership function $\mu(x)$ can be interpreted as the probability that a given element x belongs to the random set.

This interpretation of fuzzy sets as random sets has been known and used for several decades; see, e.g., [20] and references therein.

Another relationship between fuzzy and probabilities: an alternative way of gauging the degree of certainty. Another alternative – also in terms of probabilities – is to select, as $\mu(x)$, the *subjective probability* that x is possible.

In general, a subjective probability $ps(E)$ of an event E can be determined in a way which is similar to utility: as a probability p for which the lottery $L(p)$ is equivalent to the "lottery" $L(E)$ in which we have A_1 if E and A_0 otherwise. In this particular case, as the event E, we can take, e.g., the event that the majority of experts consider x to be possible.

Estimating probability that a random element belongs to a fuzzy set. In the interval case, we considered a probability distribution (uniform) on the set of all the numbers – i.e., a random number. In the interval case, the interval itself was "deterministic", so we defined the probability $P(X)$ as the probability that a random element belongs to this deterministic interval X.

A fuzzy set means, in effect, that instead of a deterministic set, we have a random set, i.e., that we have a probability measure on the class of all possible sets. Thus, it is reasonable to describe the probability $P(X)$ as the probability that a random element x belongs to the corresponding random set. Due to the formula of full probability, this probability can be described as the integral $P(X) = \int P_X(x \in S) \cdot \rho(x)\, dx$, where

- $P_X(x \in S)$ is the probability that a given element x belongs to the corresponding random set, and
- $\rho(x)$ is the probability density corresponding to randomly selecting an element.

According to the above random set interpretation of a fuzzy set, the probability $P_X(x \in S)$ that a given element x belongs to the randomly selected set is equal to the corresponding value of the membership function $\mu_X(x)$. We also know that the distribution $\rho(x)$ is uniform, so the corresponding probability density is constant: $\rho(x) = c$ for some constant c. Thus, the desired probability has the form $P(X) = c \cdot \int \mu_X(x)\, dx$.

For a "crisp" (deterministic) 1-D set, e.g. for the interval X, the integral $\int \mu_X(x)\, dx$ is simply equal to the width of the corresponding interval. Because of this, in the general 1-D case, the integral $\int \mu_X(x)\, dx$ is known in fuzzy sets as its *length* of the corresponding fuzzy set. The corresponding 2-D integral is known as an area, 3-D as a volume, and in general, as a measure of the fuzzy set etc.; see, e.g., [12]. Thus, the probability $P(X)$ is proportional to the measure of the fuzzy set X.

Similarly, the probability $P(A \cap X)$ is proportional to the measure of the intersection fuzzy set $A \cap X$, i.e.,

$$P(A \cap X) = c \cdot \int \mu_{A \cap X}(x)\, dx.$$

Therefore, the desired ratio $p_A = \dfrac{P(A \cap X)}{P(X)}$ is equal to

$$p_A = \frac{\int \mu_{A \cap X}(x)\, dx}{\int \mu_X(x)\, dx}.$$

Conclusion. Our conclusion is that

- we accept the null hypothesis for the object characterized by the fuzzy set X is the above-computed probability p_A exceeds the threshold $p^{(0)}$, and
- we reject the null hypothesis if $p_A < p^{(0)}$.

This idea can be naturally extended to fuzzy regulations. In the above text, we considered crisp regulations such as $x \le x_0$. In this case, the accept and the reject regions are crisp sets.

In practice, sometimes, we have fuzzy regulations, such as "the speed should be about 100 km/h or less". For such regulations, the accept region A is also fuzzy. The above formulas can be applied to this case as well.

ACKNOWLEDGMENTS

This work was supported in part by the research project KU 1250/4-1 "Geodätische Deformationsanalysen unter Berücksichtigung von Beobachtungsimpräzision und Objektunschärfe" funded by the German Research Foundation (DFG), by NSF grant HRD-0734825, by Texas Department of Transportation Research Project No. 0-5453, by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08, and by the Max Planck Institut für Mathematik.

The authors are thankful to the anonymous referees for valuable suggestions.

REFERENCES

[1] T. Augustin, "Neyman-Pearson testing under interval probability by globally least favorable pairs – Reviewing Huber-Strassen theory and extending it to general interval probability", *Journal of Statistical Planning and Inference*, 2002, Vol. 105, pp. 149–173.

[2] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu, *Measurement Errors in Nonlinear Models: A Modern Perspective*, Chapman & Hall/CRC, Boca Raton, Florida, 2006.

[3] R. Coppi, M. A. Gil, and H. A. L. Kiers, "The fuzzy approach to statistical analysis", *Computational Statistics and Data Analysis*, 2006, Vol. 51, pp. 1–14.

[4] T. Denoeux, M.-H. Masson, and P.-A. Hébert, "Nonparametric risk-based statistics and significance tests for fuzzy data", *Fuzzy Sets and Systems*, 2005, Vol. 153, pp. 1–28.

[5] W. A. Fuller, *Measurement Error Models*, Wiley, New York, 1987.

[6] M. A. Gil, M. López-Díaz, and D. A. Ralescu, "Overview on the development of fuzzy random variable", *Fuzzy Sets and Systems*, 2006, Vol. 157, pp. 2546–2557.

[7] L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.

[8] Interval computations website http://www.cs.utep.edu/interval-comp

[9] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control and Robotics*, Springer-Verlag, London, 2001.

[10] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

[11] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives*, John Wiley and Sons, New York, 1976.

[12] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic*, Prentice Hall, New Jersey, 1995.

[13] V. Kreinovich, H. T. Nguyen, and S. Niwitpong, "Statistical Hypothesis Testing Under Interval Uncertainty: An Overview", *International Journal of Intelligent Technologies and Applied Statistics*, 2008, Vol. 1, No. 1, pp. 1–33.

[14] H. Kutterer, "Statistical hypothesis tests in case of imprecise data", In: Proc. of the 5. Hotine-Marussi-Symposium, *International Association of Geodesy Symposia*, Springer, Berlin and New York, 2004, pp. 49-56.

[15] H. Kutterer and I. Neumann, "Multidimensional statistical tests for imprecise data", In: *Proceedings of the 6th Hotine-Marussi Symposium of Theoretical and Computational Geodesy*, Wuhan, China, 2006.

[16] R. D. Luce and H. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

[17] I. Neumann and H. Kutterer, "Congruence tests and outlier detection in deformation analysis with respect to observation imprecision", *International Journal of Applied Geodesy*, 2007, Vol. 1, No. 1, pp. 1–7.

[18] I. Neumann and H. Kutterer, "The probability of type I and type II errors in imprecise hypothesis testing", In: R. L. Muhanna and R. L. Mullen (eds.), *Proceedings of the NSF Workshop on Imprecise Probability in Engineering Analysis and Design*, Savannah, Georgia, February 2008, pp. 137–154.

[19] I. Neumann, H. Kutterer, and S. Schön, "Outlier detection in geodetic applications with respect to observation imprecision", In: R. L. Muhanna and R. L. Mullen (eds.), *Proceedings of NSF Workshop on Reliable Engineering Computing*, Savannah, Georgia, 2006, pp. 75–90.

[20] H. T. Nguyen, *An Introduction to Random Sets*, Chapman & Hall/CRC, Bocar Raton, Florida, 2006.

[21] H. T. Nguyen and V. Kreinovich, "Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications", In: R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.

[22] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 2006.

[23] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer-Verlag, New York, 2005.

[24] H. Raiffa, *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.

[25] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman & Hall/CRC, Bocar Raton, Florida, 2004.